

# **SISTEM PREDIKSI SPAM ACCOUNT PADA MEDIA SOSIAL TWITTER DENGAN MENGGUNAKAN ALGORITMA C4.5**

## **ABSTRAK**

Spam merupakan masalah yang sulit dihindari di era digital saat ini. Spam dapat digambarkan sebagai sesuatu hal yang tidak diminta dengan melakukan tindakan secara berulang yang berdampak negatif bagi pengguna lain termasuk berbagai bentuk interaksi dan perilaku akun otomatis serta upaya untuk menyesatkan atau menipu pengguna. Spam dapat berupa pesan yang dikirimkan melalui email, pesan instan, blog dan bahkan dalam media sosial. Twitter merupakan salah satu media sosial yang sangat populer dalam menyebarkan informasi maupun media diskusi. Twitter memiliki 115 juta pengguna aktif dengan 58 juta tweet yang dikirim setiap harinya. Hal ini dimanfaatkan para *spammer* untuk menjadikan twitter sebagai media untuk melakukan kegiatan *spamming*. Sebanyak 9,3% tweet yang dikirimkan di media social twitter merupakan spam. Penelitian terkait tentang mendeteksi spam di Twitter telah banyak dilakukan dan mendapatkan hasil dengan akurasi yang cukup memuaskan. Pada penelitian ini, hasil akurasi dari klasifikasi akun spam dengan menggunakan algoritma C4.5 mendapatkan akurasi model sebesar 94,42% dan akurasi klasifikasi data uji sebesar 95%.

Kata kunci : *Spam account*, Klasifikasi, *Decision Tree* , C4.5, Twitter

# **SPAM ACCOUNT PREDICTION SYSTEM IN SOCIAL MEDIA TWITTER USING C4.5 ALGORITHM**

## **ABSTRACT**

Nowadays, spam is a difficult problem to avoid . Spam can be described as something not asked and do repetitive action that negatively affects other users including various forms of interaction and behavior of automatic account as well as an attempt to mislead or deceive the user. Spam can be sent via email, instant messaging, blogs and even in social media. Twitter is one of popular social media to spreading information and media discussion. Twitter has 115 million active users with 58 million tweets sent every day. This situation is exploited by spammers to make twitter as a medium for spamming activities. 9.3% of tweet posted on the Twitter is spam. Related research on detecting spam on Twitter has been done and get accuracy result that is quite satisfactory. In this study, the results of the classification accuracy of spam accounts using the C4.5 algorithm get 94.42% model's accuracy and 95% accuracy of test data classification.

**Keywords** : : *Spam account, Classification, Decision Tree , C4.5, Twitter*